frontiers in Environmental Science

REVIEW
published: 17 June 2015
doi: 10.3389/fenvs.2015.00045

A survey of remote-sensing big data

Peng Liu *

Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China

We have entered an era of big data. It is popular to refer to the three Vs when characterizing big data: remarkable growths in the volume, velocity and variety of data. However, this statement is too general. Remote-sensing big data has several concrete and special characteristics: multi-source, multi-scale, high-dimensional, dynamic-state, isomer, and non-linear characteristics. This survey explains these characteristics in detail. Furthermore, according to whether the characteristics are closely related to the instruments or methods of data acquisition, we points out that the dynamic-state, multi-scale and non-linear characteristics are intrinsic characteristics of remote-sensing big data while the multi-source, high-dimensional and isomer characteristics are extrinsic characteristics of remote- sensing big data. In addition, we briefly review promising techniques and applications of remote-sensing big data.

Keywords: remote sensing, big data, multi-source, multi-scale, high-dimension, dynamic-state, isomer, non-linearity

OPEN ACCESS

Edited by:
Marco Casazza, "Parthenope" University of Naples, Italy

Reviewed by:
Guennady Ougolnitsky, Southern Federal University, Russia
ZhiQiang Chen, University of Missouri-Kansas City, USA

***Correspondence:**
Peng Liu, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, No. 9, Dengzhuang South Road, Beijing 100094, China
liupeng@radi.ac.cn

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *26 March 2015*
Accepted: *02 June 2015*
Published: *17 June 2015*

Citation:
Liu P (2015) A survey of remote-sensing big data. Front. Environ. Sci. 3:45. doi: 10.3389/fenvs.2015.00045

1. Introduction

Remote sensing has become one of the most important methods used to quickly and directly acquire information on the Earths surface. In recent years, with development of environmental information science, remote sensing data have played an important role in many research fields, such as atmospheric science, ecology, soil contamination, water pollution, environmental geology, environmental soil science, volcanic phenomena and evolution of the Earths crust.

The requirements of research have accelerated the development of Earth observation technologies. Many countries have rushed to launch their own satellites. **Figure 1** summarizes the number of remote sensing satellites launched by major countries in the period 1962–2014. It is seen that the USA, India and Russia are the three counties that have launched most remote sensing satellites. For most countries and regions, almost all remote sensing satellites have been launched in the period 2001–2014.

The requirements of different investigations have increased the specialization and diversity of techniques of acquiring remote sensing data. Remote sensing data often differ features in terms of their resolution, revisit cycle, spectrum, and mode of imaging. Nowadays, we can choose different remote sensing systems and datasets for different applications. A satellite can be classified as providing low-resolution imaging (e.g., MODIS[1]and Envisat[2]), mid-resolution imaging (e.g., Landsat[3], EO-1[4], Terra[5], and RADARSAT[6]), or high-resolution imaging (e.g., QuickBird[7],

[1]NASA. [Online]. Available: http://modis.gsfc.nasa.gov/about/
[2]ESA. [Online]. Available: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat
[3]USGS. [Online]. Available: https://landsat.usgs.gov
[4]NASA. [Online]. Available: https://eo1.gsfc.nasa.gov
[5]NASA. [Online]. Available: https://terra.nasa.gov/
[6]CSA. [Online]. Available: http://www.asc-csa.gc.ca/eng/satellites/radarsat2/
[7]Quickbird. [Online]. Available: http://www.satimagingcorp.com/gallery/quickbird/

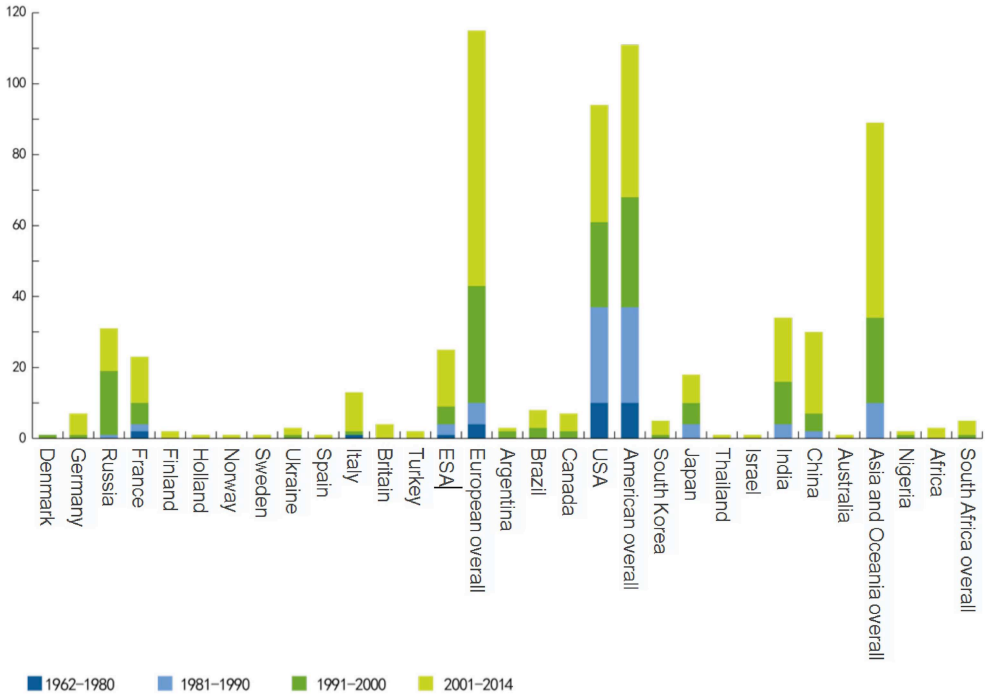

FIGURE 1 | Summary of remote sensing satellites by nation or region.

IKONOS[8], and WorldView[9]). A satellite can be classified by its mode of imaging as an optical satellite (e.g., SPOT[10], LandSat, and IKONOS), microwave satellite (e.g., TerraSAR-X[11], RADARSAT, and Envisat), or multi-mode satellite (e.g., MODIS). A satellite can be classified by its area of application as a terrestrial satellite (e.g., LandSat1-7), ocean satellite (e.g., ERS-1[12]), or meteorological satellite (e.g., MODIS). Finally, a satellite can be classified by its ability to revisit an observation area. For example, satellites of the Geostationary Operational Environmental Satellite (GOES) system can provide continuous, timely and high-quality environmental and atmospheric observations over the surface of the Earth, whereas there are also satellites with a short revisit period of 1 day (e.g., MODIS, WorldView and RapidEye[13]) and satellites with a long revisit period of 16 days (e.g., EO-1 and Landsat-7). **Table 1** gives a selection of satellites whose data are often used in environmental information science research. Overall, it is seen that there is a tremendous variety of remote sensing data.

[8]satimagingcorp. [Online]. Available: http://www.satimagingcorp.com/gallery/ikonos/

[9]digitalglobe. [Online]. Available: https://www.digitalglobe.com/about-us/content-collection

[10]airbusds. [Online]. Available: http://www.geo-airbusds.com/en/143-spot-satellite-imagery

[11]airbusds. [Online]. Available: http://www.geo-airbusds.com/terrasar-x/

[12]ESA. [Online]. Available: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/ers

[13]ESA. [Online]. Available: https://earth.esa.int/web/guest/missions/3rd-party-missions/current-missions/rapideye

Another characteristic of remote sensing data is its large volume. The volume of remote sensing data for a single scene is usually on the gigabyte level, the volume of data received by a large ground station [such as China Remote Sensing Satellite Ground Station (RSGS) in China] is usually on the terabyte level, the volume of the archive of historical data in some countries (e.g., China) is of the petabyte level, and the volume of the global archive could be of the exabyte level. Additionally, because there are so many satellites orbiting the Earth, the rate of data acquisition is very high. In the case of the RSGS, the volume of data received in 1 day exceeds 1 TB. Therefore, remote sensing data are clearly big data.

2. Features of Remote-sensing Big Data

Big data refers to a collection of data sets so large and complex that it is difficult to employ traditional data processing algorithms and models. Challenges include the acquisition, storage, searching, sharing, transfer, analysis, and visualization of the data. Scientists regularly encounter limitations due to large datasets in many areas, such as geoscience and remote sensing, complex physics simulations, and biological and environmental research. When we talk about the features of big data, it is popular to refer to the three Vs (Laney, 2001): significant growth in the volume, velocity and variety of data. However, the term the three Vs is too general. The big data of remote sensing has several concrete and special characteristics; i.e., the data

TABLE 1 | Summary of the characteristics of satellites often used in environmental information research.

Satellite	Sensor	Swath (km)	Spatial resolution (m)	Revisit capability
Airborne	Variable	Variable	>0.1	Mobilized to order
	CASI	Variable	1–2	
	Hymap	100–225	2–10	
Worldview	Panchromatic	16.4	0.46	1.1 days
	Multispectral	16.4	1.85	
Quickbird	Panchromatic	16.5	0.6	1.5–3 days
	Multispectral	16.5	2.4	
IKONOS	Panchromatic	11	1	1.5–3 days
	Multispectral	11	4	
RapidEye^	Multispectral	77 × 1500	6.5	1day
EO-1	ALI	60	30	16 days
	Hyperion	7.5	30	
Terra	ASTER	60	15,30,90	4–16 days
Terra/Aqua	MODIS	2300	250,500,1000	At least twice daily
GOES	Variable	1,4,8		Real time
ALOS	PRISM	35	4	Several times per year
SPOT-4	Panchromatic	60–80	10	11 times every 26 days
	Multispectral	60–80	20	
SPOT-5	Panchromatic	60–80	5	11 times every 26 days
	Multispectral	60-80	10	
Kompsat	Panchromatic	15	1	2–3 days
	Multispectral	15	15	
Landsat-5	TM Multispectral	185	30	Every 16 days
	TM Thermal	185	120	
Landsat-7	ETM+panchromatic	185	15	Every 16 days
	ETM+ Multispectral	185	30	
	ETM+ Thermal	185	60	
NOAA	AVHRR	2399	1100	Several times per day
Envisat	MERIS	575	300	2–3 days
Radarsat-2	Ultra-fine	20	3	Every few days
	Quad-polfine	25	8	
	Quad-pol standard	25	25	
Radarsat-1	Wide	150	30	
	Extended low	170	35	
ERS-2		100	30	35 day repeat cycle
Envisat	ASAR standard	100	30	36 day repeat cycle
	ASAR ScanSAR	405	1000	
TerraSAR-X	Spotlight	10	1	11-day repeat cycle
	Stripmap	30	3	2.5-day revisit capability
	ScanSAR	100	18	

have multi-source, multi-scale, high-dimensional, dynamic-state, isomer, and non-linearity characteristics.

The multi-source characteristic of remote-sensing big data is obvious. The fundamental reason for the multi-source characteristic is that we often use different instruments to acquire the data. Furthermore, the physical meanings of the multi-source data may be totally different. From the perspective of the imaging mechanism, the main data types are optical data, microwave data, and point cloud data. Other types of remote sensing data include stereographic pairs created from multiple photographs (often used to create three-dimensional or topographic maps) and gravity data that show the gravity situation and the amount of water available in one region. The multi-source data allows us to use and understand information from multiple viewpoints. However, they sometimes cause confusion in that we need to decide which type is the most appropriate and effective for a particular application.

Reference is often made to the multiple scales of the big data of remote sensing. The observation scale, which is also called the measurement scale, refers to the resolution, time interval, spectral range, solid angle or polarization direction (Wu and Li, 2009). Spatial scale refers to the spatial resolution and can be thought of as the size of the smallest objects that can be distinguished by sensors. A good observation often depends on the appropriate spatial scale. As a result, we have large numbers of satellites and sensors with different spatial resolutions. From the perspective of spatial resolution, there are the high-resolution satellites such as QuickBird (resolution of 0.61 m), mid-resolution satellites such as Landsat satellites (30 m), and low-resolution satellites such as MODIS (250 m). The multi-scale characteristics of the remote-sensing big data mean that it is important to select an appropriate scale and to consider scale effects in data analysis and data processing.

The high-dimensional characteristic of remote-sensing big data is mainly reflected in the spectral and temporal dimensions of the data. As examples, the AVIRIS system provides 224 spectral bands in the 0.4–2.5 μm region and MODIS with a 1-day revisit cycle provides long-time-series data. Analysis of high-dimensional image data presents both new possibilities and new challenges. High-dimensional data provide us more information about the surface of the Earth but also raise many difficulties. The first difficulty is the curse of dimensionality (Bellman, 1957). The complexity of many existing data mining algorithms is exponential with respect to the number of dimensions. With increasing dimensionality, these algorithms soon become computationally intractable and therefore inapplicable in many real applications. The second difficulty is heterogeneity. Having too few points in high-dimensional data makes efficient learning difficult in what is called the empty space phenomenon. In fact, the empty space phenomenon is a special case of the heterogeneity of big data. Apparently, high-dimensional data are far more difficult to analyze than low-dimensional data in most cases.

The big data of remote sensing always reflect a dynamic state because the Earth surface changes and the satellites move. The dynamic state of the remote-sensing big data includes both stationary parts and non-stationary parts. The changes caused by the Earth orbiting the Sun and rotating about its own axis (e.g., the alternation of seasons and climate changes) are stationary from the point of view of a stochastic process. Changes caused by human activities and natural disasters, such as the evolution of a city and volcanic eruptions and earthquakes, are non-stationary stochastic processes. The stationary features of the remote sensing big data show us the explicit law that is easily represented by statistical method. However, the non-stationary feature increases the difficulty of analysis of the big data. Methods that are more advanced are required to find the implicit law hidden within the remote-sensing big data.

The isomer characteristic of remote-sensing big data often refers to different data representation structures for the same geographic coordinates. The most obvious isomer data are raster data and vector data. The raster data type consists of rows and columns of cells. Each cell stores a single value. Raster data can be images comprising individual pixels. Vector data express geographical features and geometrical shapes as vectors. They are often used by the Geographic Information System (GIS), and many are derived from remote-sensing raster data. Raster data such as an optical image are usually stored as matrix data structures on a computer. Vector data, however, are more complicated and are stored in a variety of data structures, such as linked lists, trees, and graphs. In some cases, one type of structured data can partly transform into another type of structured data. As an example, after we extract road and building information from raster data, we often represent the extracted information by vector data. When the isomer characteristic of the data is considered, it is not easy to explore the relationships between the different types of data, although isomer data do provide us more information than one single type of data. For example, the registration between image data and a vector map is usually far more difficult than the band registration of multi-source optical images. Overall, the variety of data challenges the processing and management of the isomer data.

When we take the Earth and our natural environments as systems, they always have non-linear characteristics. As result, Earth observation data acquired employing remote sensing methods have nonlinear characteristics. For example, time series of remote sensing data are typically nonlinear and noisy. Such time series usually cannot be studied satisfactorily by linear time series analysis. Although traditional linear techniques are useful for studying characteristic oscillations in detail, these methods fail to detect any non-linear correlations present and cannot provide a complete characterization of the underlying dynamics. Therefore, we need advanced non-linear analysis methods that are suited to the characterization of the dynamics in a noisy, high-dimensional and under-determined system. Furthermore, only the successfully characterization of irregular time series from mathematical models based on non-linearity will allow us gain an insight into the nature of remote-sensing big data.

It is important that we consider the multi-source, multi-scale, high-dimensional, dynamic-state, isomer, and non-linear characteristics of remote sensing data when using remote sensing to understand geo-processes, and the characteristics are fundamental assumptions and priors when we analyze remote-sensing big data and extract information from the data.

3. Typical Applications

There are already many applications of remote-sensing big data. For example, Google Maps[14] and Google Earth[15] have for years been two of the most popular Internet mapping applications. Both Google Earth and Google Maps access the Google Earth Engine, a platform providing an extremely large repository of geo-referenced satellite imagery, terrain data, and vector data (such as borders, roads, population centers, soil information and climate information). In time-series analysis, a new algorithm (Zhu et al., 2015) that generates synthetic LandSat images based on all available LandSat data has been developed. The algorithm has provided promising results in filling gaps and removing cloud, shadow and snow. In research on environmental systems, remote-sensing big data are playing a key role in providing accurate estimates of surface fluxes of greenhouse gases with accurate estimates of associated uncertainties at intermediate spatiotemporal scales (Miller et al., 2014; Zscheischler et al., 2014). In machine learning, a deep architecture that is capable of learning feature representations from both labeled and unlabeled data has attracted the attention of many researchers (Hinton and Salakhutdinov, 2006). The architecture incorporates both unsupervised pre-training and supervised fine-tuning strategies to construct models (Bengio et al., 2007); unsupervised stages learn data distributions without using label information and supervised stages perform a local search for fine tuning. Deep learning is also applied in many remote-sensing data analyses (Han et al., 2015; Tang et al., 2015). Employing remote-sensing big data, a Hessian-based method (Kalmikov and Heimbach, 2014) has been successfully applied to uncertainty quantification in estimation of the global ocean state. Data assimilation technologies that are often closely related to remote-sensing big data are developing rapidly in the research areas of the ocean (Kalmikov and Heimbach, 2014; Coelho et al., 2015), hydrology (Panzeri et al., 2015; Yucel et al., 2015), atmosphere (Barcons et al., 2015), soil (Liang et al., 2015), and agriculture (Huang et al., 2015). Many of these applications consider the multi-source, multi-scale, high-dimensional, dynamic-state, isomer, and non-linear characteristics of remote-sensing big data.

4. Discussion

As discussed above, remote-sensing big data have multi-source, multi-scale, high-dimensional, dynamic-state, isomer, and non-linear characteristics. These characteristics arise from different genetics. Among the characteristics, the dynamic-state, multi-scale and non-linear characteristics are not due to the sensors or instruments. Even if we do not use remote sensing to acquire the data, natural phenomena always have dynamic-state, multi-scale and non-linear characteristics. These characteristics have little to do with the method of acquisition or the hardware of sensors. However, the multi-source, high-dimensional and isomer characteristics are closely related to the sensors or instruments. For example, if we use multi-spectral sensors, the number of dimensions of the data will be not as high as that if we use hyperspectral sensors. Furthermore, multi-source and isomer characteristics refer to the different aspects or appearances of the same object when using different measurement instruments or representative structures. Additionally, the isomer characteristic depends more on the external form of the data structure in the computing process. We thus refer to the dynamic-state, multi-scale and non-linear characteristics as the intrinsic characteristics of remote-sensing big data, and the multi-source, high-dimensional and isomer characteristics as the extrinsic characteristics of remote-sensing big data.

There is no doubt that existing techniques and methods are too limited to solve all the problems of remote-sensing big data completely. Fortunately, we are witnessing technological leapfrogging. In the case of the high-dimensional characteristic, there are many dimension-reduction methods (such as manifold learning, Talwalkar et al., 2013) that could be applied to big data. For the multi-source problem, data fusion technology is promising. Furthermore, transfer learning (Pan and Yang, 2010) could be used to explore the isomers big data. Considering the dynamic state, techniques developed in many studies on online-learning and active learning conceptions (Ruiz et al., 2014) could be used to analyze remote-sensing big data. In addition, in recent years, popular research on sparse representation has already provided promising results for disposing remote-sensing dynamic data sets (Wang et al., 2014). To address the non-linear characteristic, over the last two decades, many non-linear time series methods have been developed in the theory of non-linear dynamics, commonly known as chaos theory (Lorenz, 2005). We are already surrounded by remote-sensing big data, and challenges are always followed by opportunities. Remote-sensing big data are providing scientists with a new way of understanding the external world.

[14] googlemaps. [Online]. Available: https://maps.google.com
[15] googleearth. [Online]. Available: https://earth.google.com

Acknowledgments

It is supported by: NSFC(41471368).

References

Barcons, J., Folchb, A., Afifa, A. S., and Miró, J. R. (2015). Assimilation of surface AWS using 3DVAR and LAPS and their effects on short-term high-resolution weather forecasts. *Atmos. Res.* 156, 160–173. doi: 10.1016/j.atmosres.2014.12.019

Bellman, R. E. (1957). *Dynamic Programming, 1st Edn.* Princeton, NJ: Princeton University Press.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2007). Greedy layer-wise training of deep networks. *Adv. Neural Inf. Process. Syst.* 19, 153–160. doi: 10.1.1.70.2022

Coelho, E. F., Hogan, P., Jacobs, G., Thoppil, P., Huntley, H. S., Haus, B. K., et al. (2015). Ocean current estimation using a multi-model ensemble kalman filter during the grand lagrangian deployment experiment (glad). *Ocean Model.* 87, 86–106. doi: 10.1016/j.ocemod.2014.11.001

Han, J., Zhang, D., Cheng, G., Guo, L., and Ren, J. (2015). Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Trans. Geosci. Remote Sens.* 53, 3325–3337. doi: 10.1109/TGRS.2014.2374218

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., et al. (2015). Improving winter wheat yield estimation by assimilation of the leaf area index from landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121. doi: 10.1016/j.agrformet.2015.02.001

Kalmikov, A. G., and Heimbach, P. (2014). A hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM J. Sci. Comput.* 36, 267–295. doi: 10.1137/130925311

Laney, D. (2001). *3d Data Management: Controlling Data Volume, Velocity and Variety*. Stamford, CT: META Group Inc.

Liang, J., Li, D., Shi, Z., Tiedje, J. M., Zhou, J., Schuur, E. A., et al. (2015). Methods for estimating temperature sensitivity of soil organic matter based on incubation data: a comparative evaluation. *Soil Biol. Biochem.* 80, 127–135. doi: 10.1016/j.soilbio.2014.10.005

Lorenz, E. (2005). Designing chaotic models. *J. Atmos. Sci.* 62, 1574C–1587C. doi: 10.1175/jas3430.1

Miller, S. M., Michalak, A. M., and Wofsy, S. C. (2014). Reply to Hristov et al.: linking methane emissions inventories with atmospheric observations. *Proc. Natl. Acad. Sci. U.S.A.* 111:E1321. doi: 10.1073/pnas.1401703111

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Panzeri, M., Riva, M., Guadagnini, A., and Neuman, S. (2015). Enkf coupled with groundwater flow moment equations applied to lauswiesen aquifer, germany. *J. Hydrol.* 521, 205–216. doi: 10.1016/j.jhydrol.2014.11.057

Ruiz, P., Mateos, J., Camps-Valls, G., Molina, R., and Katsaggelos, A. K. (2014). Bayesian active remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* 52, 2186–2196. doi: 10.1109/TGRS.2013.2258468

Talwalkar, A., Kumar, S., Mohri, M., and Rowley, H. A. (2013). Large-scale SVD and manifold learning. *J. Mach. Learn. Res.* 14, 3129–3152.

Tang, J., Deng, C., Huang, G., and Zhao, B. (2015). Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE Trans. Geosci. Remote Sens.* 53, 1174–1185. doi: 10.1109/TGRS.2014.2335751

Wang, L., Lu, K., Liu, P., Ranjan, R., and Chen, L. (2014). IK-SVD: dictionary learning for spatial big data via incremental atom update. *Comput. Sci. Eng.* 16, 41–52. doi: 10.1109/MCSE.2014.52

Wu, H., and Li, Z.-L. (2009). Scale issues in remote sensing: a review on analysis, processing and modeling. *Sensors* 9, 1768–1793. doi: 10.3390/s903 01768

Yucel, I., Onen, A., Yilmaz, K., and Gochis, D. (2015). Calibration and evaluation of a flood forecasting system: utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *J. Hydrol.* 523, 49–66. doi: 10.1016/j.jhydrol.2015.01.042

Zhu, Z., Woodcock, C. E., Holden, C., and Yang, Z. (2015). Generating synthetic landsat images based on all available landsat data: Predicting landsat surface reflectance at any given time. *Remote Sens. Environ.* 162, 67–83. doi: 10.1016/j.rse.2015.02.009

Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., et al. (2014). Impact of large-scale climate extremes on biospheric carbon fluxes: an intercomparison based on MsTMIP data. *Glob. Biogeochem. Cycles* 28, 585–600. doi: 10.1002/2014GB 004826

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.